

# **ELPOD: Nástroj pro analýzu výsledků vyhledávání na webu**

## **ELPOD: Tools for Analysis of Web Search Engine Results**

## Zadání bakalářské práce

Student:

**Vojtěch Garba**

Studijní program:

B2647 Informační a komunikační technologie

Studijní obor:

2612R025 Informatika a výpočetní technika

Téma:

**ELPOD: Nástroj pro analýzu výsledků vyhledávání na webu**  
**ELPOD: Tools for Analysis of Web Search Engine Results**

Zásady pro vypracování:

Cílem této práce je vytvoření nástroje pro sledování pozice v internetových vyhledávačích, a to jak placených reklamních kampaní, tak běžného vyhledávání.

1. Seznamte se s problematikou vyhledávání, princip hodnocení klíčových slov, četnosti jejich vyhledávání a podobně.
2. Seznamte se s existujícími aplikacemi sledující pozice ve vyhledávačích SEO Rank Tracker a WebCEO, popište jejich funkcionalitu a možnosti srovnávacích testů konkurenčních webů.
3. Seznamte se s problematikou načítání a parsování výsledku vyhledávání pro služby Google, Binga, Seznam. Oddělte výsledky běžného vyhledávání od komerčních sdělení.
4. Proveďte analýzu, návrh a implementaci aplikace, která bude řešit problematiku dlouhodobého sledování pozic ve vyhledávačích, bude sledovat vývoj konkurenčních webů a potenciálně vznikajících konkurentů.
5. Implementaci proveďte na platformě .NET jako kombinaci webové a desktopové aplikace.
6. Zaměřte se zejména na problematiku reduplikace vyhledávání a optimalizace při dotazování.
7. Výsledky sledování statistik vhodně prezentujte v podobě grafů a tabulek.
8. Navrhněte další možnosti zlepšení a rozšíření této problematiky.
9. Srovnajte výslednou aplikaci s existujícími systémy.

Seznam doporučené odborné literatury:

Podle pokynů vedoucího bakalářské práce. Nastudujte projekty SEO Rand Tracker, WebCEO a další.

Formální náležitosti a rozsah bakalářské práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí bakalářské práce: **Ing. Radoslav Fasuga, Ph.D.**

Datum zadání: 16.11.2012

Datum odevzdání: 07.05.2013



doc. Dr. Ing. Eduard Sojka  
vedoucí katedry

prof. RNDr. Václav Snášel, CSc.  
děkan fakulty

Prohlašuji, že jsem tuto bakalářskou práci vypracoval samostatně. Uvedl jsem všechny literární prameny a publikace, ze kterých jsem čerpal.

V Ostravě 2. května 2013

  
.....

Tímto bych chtěl poděkovat mému vedoucímu bakalářské práce Ing. Radoslavovi Fasugovi, Ph.D., za velmi motivující a dobré vedení, jež dalo vzniku této práce.

## **Abstrakt**

Tato práce popisuje vývoj aplikace, která slouží pro sledování pozic v internetových vyhledávačích a zabývá se nejen přirozeným vyhledáváním, ale také placenými kampaněmi. V první části práce je popsána obecná teorie. Co je to optimalizace pro vyhledávače, jak vyhledávače fungují, hodnotí webové stránky a sestavují výsledky vyhledávání. Další část práce popisuje několik v již existujících aplikací zabývajících se touto problematikou s ohledem na funkčnost a omezení neplacených verzí. Předposlední část popisuje samotný vývoj aplikace a dosažené výsledky. Nakonec je aplikace srovnána s existujícími systémy a popsány možnosti vylepšení.

**Klíčová slova:** Vyhledávač, Google, Seznam, Bing, pozice ve vyhledávači, SEO, SEM, Rank Tracker, Web Ceo

## **Abstract**

This thesis describes the development of an application that is used for monitoring positions in search engines and discusses not only the natural search, but also paid campaigns. The first part describes the general theory. What is search engine optimization, how search engines work, evaluate websites and compile search results. The next part describes several existing applications that solving this issue with regard to the functionality and limitations of unpaid versions. The next part describes the development of applications and achieved results. Finally, the application is compared with existing systems and describes the options for enhancement.

**Keywords:** Search engine, Google, Seznam, Bing, position in search engine, SEO, SEM, Rank Tracker, WEB Ceo

## **Seznam použitých zkratek a symbolů**

API	– Application Programming Interface (rozhraní pro programování aplikací)
URL	– Uniform Resource Locator (jednotný lokátor zdrojů)
PPC	– Pay Per Click (platba za klik)
SŘBD	– Systém Řízení Báze Dat

## Obsah

<b>1</b>	<b>Úvod</b>	<b>5</b>
<b>2</b>	<b>ELPOD</b>	<b>6</b>
2.1	SEM . . . . .	6
2.2	SEO . . . . .	6
<b>3</b>	<b>Vyhledávače</b>	<b>8</b>
3.1	Fulltextový vyhledávač . . . . .	8
3.2	Katalog . . . . .	8
3.3	Jak fungují fulltextové vyhledávače? . . . . .	8
3.4	Podíl vyhledávačů . . . . .	10
3.5	Známkování podle vyhledávačů . . . . .	11
3.6	Sestavení výsledků vyhledávání . . . . .	14
3.7	Klíčová slova . . . . .	16
<b>4</b>	<b>Existující aplikace pro sledování pozic ve vyhledávači</b>	<b>18</b>
4.1	Přehled . . . . .	18
4.2	SEO PowerSuite . . . . .	18
4.3	Web CEO . . . . .	19
4.4	Seo Servis . . . . .	21
4.5	Seo nástroje . . . . .	21
<b>5</b>	<b>Vlastní nástroj</b>	<b>23</b>
5.1	Vize . . . . .	23
5.2	Specifikace požadavků . . . . .	23
5.3	Okolí aplikace . . . . .	24
5.4	Diagram případu užití . . . . .	24
5.5	Stahování výsledků vyhledávání . . . . .	25
5.6	Zobrazení výsledků vyhledávání . . . . .	28
5.7	Komunikace s webovou aplikací . . . . .	29
5.8	Výsledná aplikace . . . . .	30
<b>6</b>	<b>Srovnání s existujícími systémy</b>	<b>33</b>
<b>7</b>	<b>Závěr</b>	<b>35</b>
<b>8</b>	<b>Reference</b>	<b>36</b>
	<b>Přílohy</b>	<b>36</b>
<b>A</b>	<b>Obsah DVD</b>	<b>37</b>
<b>B</b>	<b>Datový Slovník</b>	<b>38</b>

## Seznam tabulek

1	Podíl vyhledávačů v ČR . . . . .	10
2	Podíl vyhledávačů ve světě . . . . .	11
3	Ukázka datového slovníku, tabulka storedResults . . . . .	28
4	Omezení aplikací . . . . .	34
5	Obsah DVD . . . . .	37



## Seznam obrázků

1	Cesta robota . . . . .	9
2	Podíl vyhledávačů v ČR . . . . .	11
3	Podíl vyhledávačů ve světě . . . . .	12
4	Předávání PageRanku mezi stránkami . . . . .	13
5	Ukázka výsledků vyhledávání - Seznam . . . . .	15
6	Okolí aplikace . . . . .	24
7	Diagram případu užití . . . . .	25
8	Sekvenční diagram - sestavení přehledu . . . . .	29
9	Ukázka aplikace - záložka Zobrazení . . . . .	30
10	Ukázka aplikace - záložka Stažení . . . . .	31
11	Ukázka aplikace - záložka nastavení . . . . .	32

## Seznam výpisů zdrojového kódu

1	Stahování a parsování dat . . . . .	26
2	SQL dotaz pro výběr dat . . . . .	28

## 1 Úvod

Dnešní doba je doba internetová a kdokoliv, kdo nechce zůstat pozadu a před svými zákazníky být prakticky neviditelný, musí mít svou webovou prezentaci. Těch ovšem neustále přibývá a jen "být na internetu" přestává stačit a pak je potřeba začít s nějakou vhodnou optimalizací. Nejlepší bude optimalizace stránek pro vyhledávače, neboli SEO. Ale ani tím to nekončí. Techniky se vyvíjejí, obsah se aktualizuje a konkurence nespí.

Dokonce i po dosažení výborných příček na pozicích ve vyhledávačích nelze usnout na vavřínech. Nadále musíme sledovat vývoj nejen vlastních webových stránek, ale i těch konkurenčních. Na internetu existuje spousta nástrojů, zabývajících se problematikou sledování pozic ve vyhledávačích. Ať už se jedná o placené produkty, nebo volně šiřitelné aplikace.

Hlavním cílem této práce, jak je patrné z jejího názvu, je vývoj nástroje zabývajícího se stahováním a analýzou výsledků vyhledávání. Nejprve se však musíme teoreticky seznámit s problematikou okolo funkčnosti vyhledávačů a způsob jakým tyto výsledky vyhledávání sestavují. Každý z nich pracuje odlišně a také stránky hodnotí jinak. Poté si představíme několik existujících aplikací, které se touto problematikou zabývají. Popíšeme si jejich funkce, výhody a nevýhody a pokusíme se co nejlépe pochopit jejich funkčnost. Poté můžeme opravdu začít s vývojem aplikace.

Chceme-li to probrat podrobněji, aplikace by měla být schopna stahovat a ukládat pro jakákoliv klíčová slova určitý počet výsledků vyhledávání a zpětně je zobrazovat příslušnou formou. Samotné přirozené vyhledávání by však nestačilo a proto přidáme i sledování placených kampaní. Konečným porovnáním s ostatními aplikacemi zhodnotíme dosažené výsledky.

## 2 ELPOD

Elektronické podnikání, neboli ve zkratce ELPOD je v poslední době velice rozmáhajícím se trendem. Firem, které představují a nabízejí své produkty na internetu, každým dnem přibývá a konkurence roste. Ať už se jedná o internetové obchody, slevové portály, nebo služby, je zde široká a rozmanitá nabídka. Kategorie se nevztahuje pouze k informačním a komunikačním technologiím, ale zahrnuje prakticky cokoli. Na internetu tak pořídíte vše možné od hraček pro děti, přes vybavení domácnosti nábytkem až po služby detektivních kanceláří. Jedním ze způsobů jak docílit velké návštěvnosti a tím i dostatek potencionálních zákazníků je získání prvních příček na pozicích ve vyhledávačích.

### 2.1 SEM

*Search Engine Marketing* (SEM), neboli marketing ve vyhledávačích vychází z předpokladu, že nejdůležitějším faktorem je být vyhledávačem nalezen a následně zobrazen uživateli, na základě klíčového slova, které souvisí s obsahem stránky.<sup>1</sup>

SEM nevyužívá klasickou reklamní strategii tzv. push marketing, která tlačí produkty či služby k zákazníkům pomocí billboardů, reklamních letáků, rádiového a televizního vysílání. Zákazník často o takovou reklamu nemá zájem a prostě ji ignoruje. Naopak SEM nabízí zákazníkům to, o co zájem mají, to co právě vyhledávají, nebo potřebují. Nejlepší situace nastává ve chvíli, kdy se zobrazíte zákazníkovi, který aktivně vyhledává produkty či služby, Vámi nabízené.

Základní rozdělení tkví ve dvou hlavních částech. První metoda spolupráce s vyhledávači je placená reklama. Jednoduše si koupíte nejlukrativnější pozice ve vyhledávačích.

Několik PPC služeb:

- Google AdWords
- Google AdSense
- Sklik.cz
- AdFox.cz
- Etarget.cz

V druhé části se snažíme stránky co nejvíce přizpůsobit tak, aby byly pro vyhledávač co nejrelevantnější a při zadání návštěvníkova dotazu umístil tyto stránky na co nejlepší pozici. K tomuto je nejlepší použít SEO.

### 2.2 SEO

Zkratka SEO pochází z anglické fráze *Search Engine Optimalization*, tedy přeloženo do češtiny optimalizace stránek pro vyhledávače. Často se vyskytuje nesprávné spojení

<sup>1</sup>KUBÍČEK, Michal. *Velký průvodce SEO: jak dosáhnout nejlepších pozic ve vyhledávačích*. Vyd. 1. Brno: Computer Press, 2008, 318 s. ISBN 978-80-251-2195-5. (dále jen Kubiček)

SEO optimalizace, což je velice nepřesné, jelikož následný překlad by vypadal takto: „optimalizace optimalizace stránek pro vyhledávače“.

Velice jednoduše řečeno se jedná o soubor technik, které zajišťují webovým stránkám co nejlepší pozici ve vyhledávačích a to právě optimalizací, neboli přizpůsobením stránek tak, aby byly vyhledávači co nejlépe ohodnoceny a jednoduše nalezeny.<sup>2</sup>

Několik důležitých SEO technik:

- **Web copywriting** popisuje správný způsob psaní textu na webových stránkách, který by měl obsahovat správné množství klíčových slov a měl by být unikátní. To znamená, že každý text na různých stránkách by měl být odlišný, nedoporučují se ani odstavce s pozměněným pořadím vět, vyhledávač může vyhodnotit text jako duplicitní a relevance je poté přikládána jen jednomu z nich. Text by neměl být zastaralý a měl by být pravidelně aktualizován.
- **Krátké a srozumitelné URL adresy**, které by neměly obsahovat nahodilý shluk znaků a čísel, naopak je vhodné, aby v sobě měly zakomponováno klíčové slovo.
- **Validní HTML kód** a správné použití značek v souladu s konvencemi s důrazem na značky h1 až h6 sloužící pro nadpisy.
- **Linkbuilding** je technika zajišťující zpětné odkazy na webovou stránku. Odkazy se registrují do katalogů, vkládají do internetových fór apod. Vyhledávač vyhodnocuje tyto odkazy směřující z různých webových stránek a přikládá jim určitou váhu, která je závislá na pageranku stránky, na které se odkaz nachází.

---

<sup>2</sup>KUBÍČEK, Michal a Jan LINHART. 333 tipů a triků pro SEO: [sbírka nejlepších technik optimalizace webů pro vyhledávače]. Vyd. 1. Brno: Computer Press, 2010, 262 s. ISBN 978-80-251-2468-0.(dále jen Kubíček a Linhart)

## 3 Vyhledávače

### 3.1 Fulltextový vyhledávač

Jedná se o službu, která se pokouší uživateli dle zadaného klíčového slova poskytnout co možná nejrelevantnější výsledky vyhledávání, což mohou být webové stránky, obrázky, dokumenty i multimediální obsah. Tyto vyhledávače aktivně prohledávají všechny webové stránky v celé síti internet a vytváří si vlastní databázi, do které tyto stránky indexují (stahují a ukládají). Takto, mimo jiné, prohledávají i katalogy. Indexace neprobíhá jen u nově nalezených stránek, nýbrž i u stávajících, vyhledávači již známých stránek a to i několikrát denně, informace již uložené v databázi vyhledávače jsou takto aktualizovány.

### 3.2 Katalog

V podstatě se jedná o webovou stránku rozdělenou podle kategorií, která shromažďuje odkazy na ostatní stránky. I zde jsou podporovány vyhledávací funkce. Katalogový vyhledávač zobrazuje výsledky nalezené v rámci katalogové databáze, pokud byla nalezena shoda s hledaným výrazem v titulku, či popisu a tento záznam byl zařazen v příslušné kategorii.<sup>3</sup>

Toto je hlavní rozdíl oproti fulltextovým vyhledávačům. Odkazy do databází katalogů jsou přidávány manuálně a vyhledávací nástroj katalogu nabídne uživateli výsledky vyhledávání pouze z vlastní databáze ručně přidaných, neboli zaregistrovaných odkazů. Mezi katalogy například patří: `firmy.cz/`, `najisto.centrum.cz/`, nebo `najduvse.cz/`.

### 3.3 Jak fungují fulltextové vyhledávače?

Jak bylo již dříve zmíněno, vyhledávače procházejí webové stránky v celé síti internet a stahují je do své databáze. Toto zajišťují speciální programy, které jsou neustále vysílány po síti a jejich úkolem je právě procházení a stahování stránek do centrální databáze vyhledávače. Každý z těchto programů bývá nazýván robot, pavouk či crawler a dokáže procházet a stahovat desítky stránek za sekundu. Jejich práce však není nějak chaotická a stránky procházejí dle algoritmu, který určuje, jaké webové prezentace budou procházeny a kolik jednotlivých stránek této prezentace se projde.<sup>4</sup>

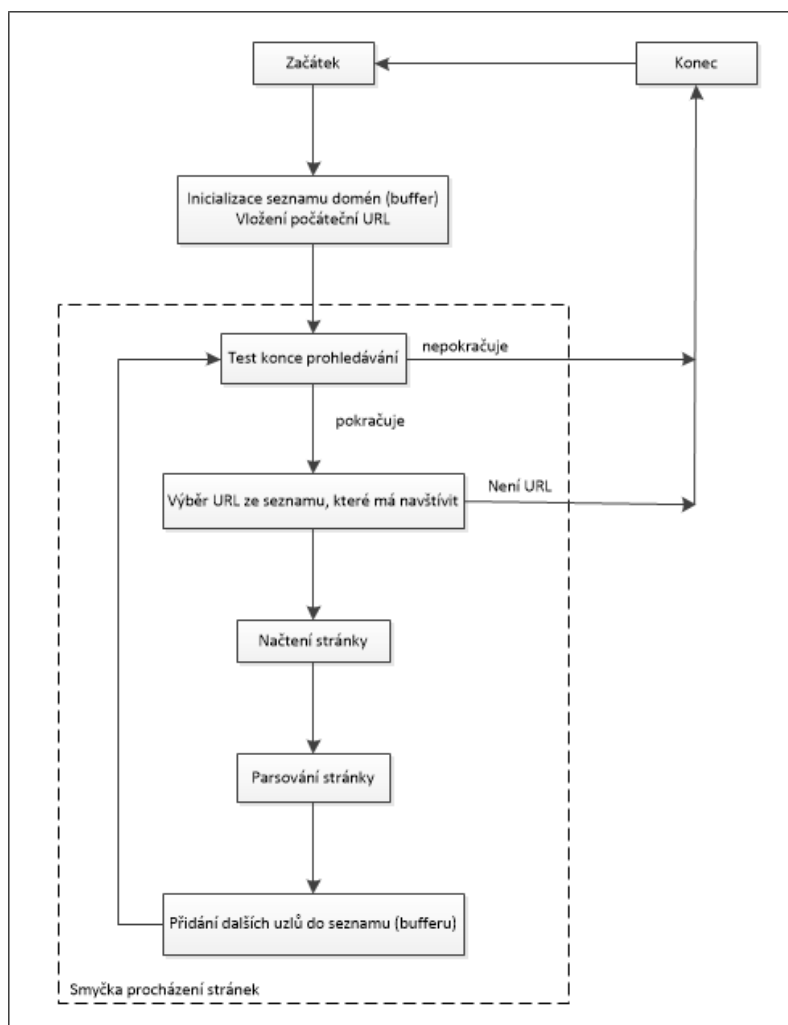
#### 3.3.1 Práce robotů

Robot začíná tím, že si z vlastní databáze vybere odkaz na stránku, kterou chce navštívit. Až tam dorazí, tak si stránku přečte a její obsah uloží do databáze vyhledávače. Následně pak ve zdrojovém kódu pátrá po párových značkách `<a>` a `</a>` v nichž jsou odkazy na další stránky. Když takovýto odkaz nalezne, přejde na novou stránku a proces se opakuje. V případě že žádný takový odkaz nenalezne, vrátí se o úroveň zpět a celý proces

<sup>3</sup>Kubíček, s. 29.

<sup>4</sup>Kubíček, s. 35.

pokračuje. Je to umožněno rekurzivním algoritmem, který mu dovoluje volat sám sebe. Díky tomu tak může okamžitě reagovat na změny stránek a nikdy nebude procházet stejnou cestu. Cesta robotů je znázorněna na obrázku číslo 1.<sup>5</sup>



Obrázek 1: Cesta robotů

Existují různé druhy robotů, přičemž každý z nich se specializuje na něco jiného<sup>6</sup>:

- Stahovač do databáze ukládá výtahy z dokumentů. Za pomoci indexů definuje vztahy klíčových slov nebo frází nalezených na webové stránce k její URL adrese.
- Prohlížeč, jehož úkolem je sledovat okolí stránek a vyhledávat například názvy obrázků, jejichž náhledy ukládá do databáze spojené s odkazem na webovou stránku,

<sup>5</sup>Kubíček, s. 36

<sup>6</sup>Kubíček, s. 36.

kde se nacházejí. Jejich funkce není příliš odlišná od stahovačů, pouze se specializují na obrázky. Obdobně pak existují podobné typy robotů zaměřující se na další typy dokumentů jako například PDF, DOC, DOCx a podobně.

- Kontrolor hledá odkazy na neexistující webové stránky. Pokud robot opakovaně stránku nenalezá, předá pokyn k jejímu smazání z hlavní databáze a tím i z výsledku vyhledávání.
- Statistici a počtáři střežadí informace o počtu stránek, odkazů a míry využití jednotlivých webových stránek z čehož poté vyplývá, které stránky jsou oblíbenější a které nikoliv.

### 3.4 Podíl vyhledávačů

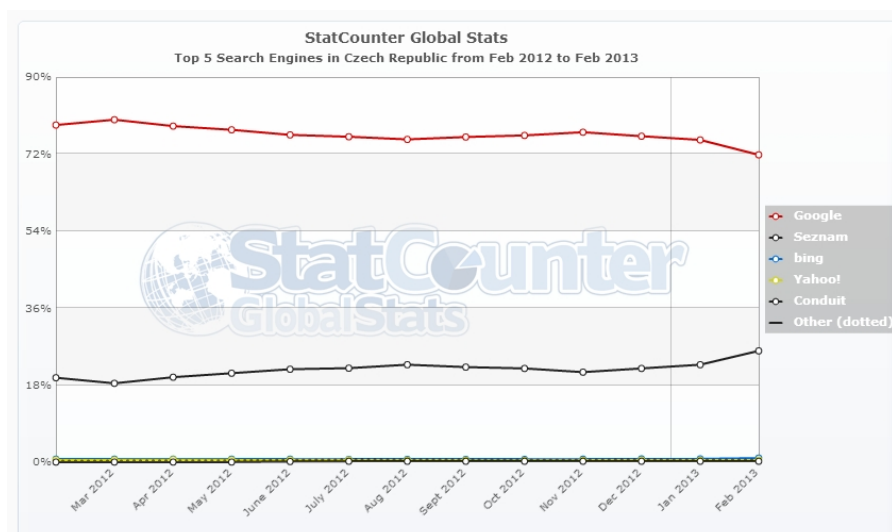
Dříve byl vývoj podílů využívání vyhledávačů velice dynamický a změny v bojích o procenta podílů byly dramatické. Dnes už jsou však jejich pozice rozděleny a během posledního roku nedochází k žádným razantním změnám. Mezi trojici nejpoužívanějších vyhledávačů ve světě patří Google, Bing a Yahoo. V České republice je to pak Google, Seznam a Bing. V obou těchto oblastech však nejvíce dominuje Google. Za poslední rok ve světě používalo Google přibližně 91% uživatelů, v České republice to bylo necelých 77%. Druhou příčku ve světě obsadil Bing s ročním průměrem pouze 3,3% a za ním Yahoo s pouhými 3%. V České republice se jako druhý umístil Seznam s ročním průměrem přibližně 22% a jako třetí Bing s 0,7%.<sup>7</sup> Jednotlivé podíly lze vidět na obrázcích 2 a 3, přesná čísla pak v tabulkách 1 a 2.

Datum	Google	Seznam	Bing	Yahoo!	Conduit	Other
2012-02	78,79	19,7	0,7	0,48	0	0,32
2012-03	80,07	18,42	0,68	0,51	0	0,32
2012-04	78,54	19,84	0,68	0,54	0	0,39
2012-05	77,69	20,79	0,65	0,49	0	0,37
2012-06	76,5	21,72	0,69	0,49	0,09	0,52
2012-07	76,07	21,97	0,62	0,5	0,16	0,68
2012-08	75,41	22,73	0,62	0,46	0,16	0,62
2012-09	75,98	22,21	0,64	0,41	0,16	0,6
2012-10	76,4	21,87	0,63	0,39	0,17	0,53
2012-11	77,1	21,03	0,68	0,44	0,18	0,57
2012-12	76,17	21,87	0,71	0,47	0,18	0,6
2013-01	75,32	22,76	0,75	0,43	0,17	0,56
2013-02	71,81	26,02	0,92	0,51	0,17	0,57

Tabulka 1: Podíl vyhledávačů v ČR

<sup>7</sup>STATCOUNTER. *GlobalStats* [online]. 2013 [cit. 2013-03-26]. Dostupné z: <http://gs.statcounter.com/>





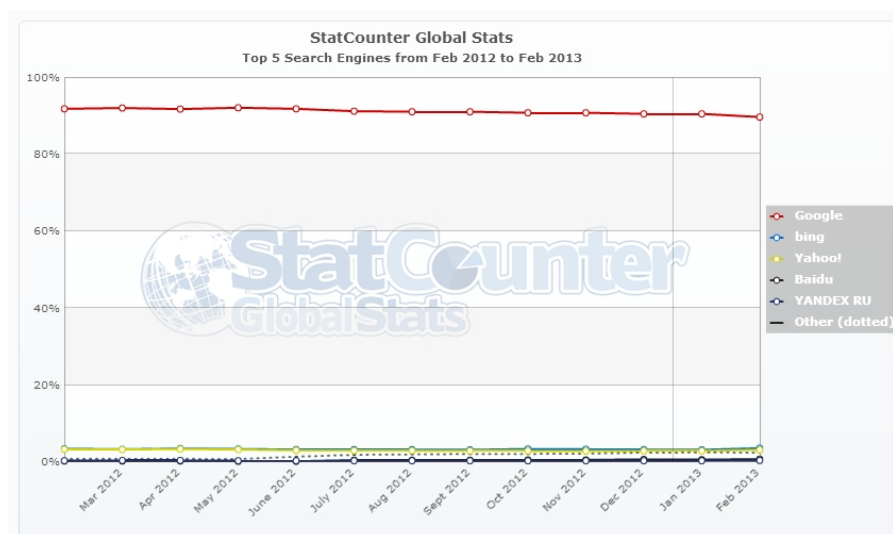
Obrázek 2: Podíl vyhledávačů v ČR

Datum	Google	Bing	Yahoo!	Baidu	YANDEX	Other
2012-02	91,78	3,46	3,31	0,26	0,36	0,82
2012-03	92,01	3,33	3,24	0,25	0,38	0,79
2012-04	91,7	3,5	3,36	0,26	0,38	0,8
2012-05	92,04	3,36	3,26	0,22	0,37	0,75
2012-06	91,75	3,27	3,04	0,23	0,35	1,36
2012-07	91,17	3,22	2,95	0,45	0,34	1,87
2012-08	91,01	3,22	2,98	0,5	0,33	1,97
2012-09	91,04	3,16	2,91	0,49	0,33	2,07
2012-10	90,75	3,35	2,91	0,54	0,36	2,09
2012-11	90,75	3,32	2,84	0,58	0,35	2,17
2012-12	90,43	3,26	2,89	0,66	0,35	2,41
2013-01	90,47	3,19	2,88	0,63	0,35	2,47
2013-02	89,64	3,62	3,17	0,73	0,42	2,42

Tabulka 2: Podíl vyhledávačů ve světě

### 3.5 Známkování podle vyhledávačů

Každý vyhledávač určuje pozici webových stránek ve vyhledávání dle vlastního hodnotícího systému - jedná se o tzv. *ranky*. Stránky jsou hodnoceny dle různých kritérií, například podle počtu zpětných odkazů, podle významnosti stránky, dle počtu zaindexovaných stránek, nebo dokonce podle počtu archivovaných stránek.



Obrázek 3: Podíl vyhledávačů ve světě

### 3.5.1 PageRank (PR)- hodnocení dle Google

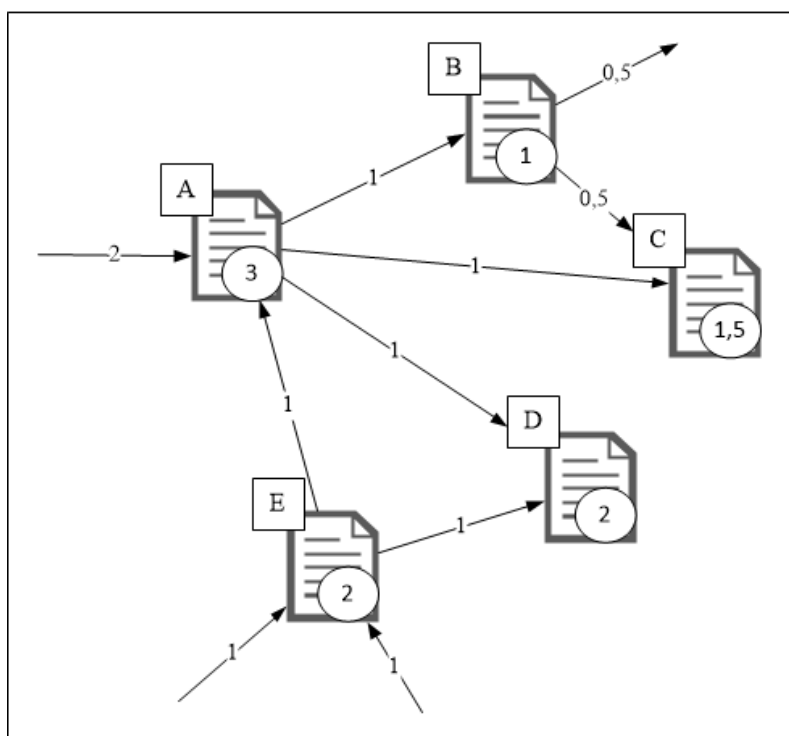
Jedná se o algoritmus, který na stupnici od 0 do 10 ukazuje relevantnost webové stránky, o jehož vznik se zasloužili dva zakladatelé společnosti Google, Lawrence Page, podle něhož byl PageRank pojmenován a Sergey Brin. Algoritmus funguje na principu Kandall-Weiovy teorie hodnocení, která porovnává význam lidí a věcí na základě vlivu, který na sebe vzájemně mají.<sup>8</sup>

V jednoduchosti to funguje tak, že PageRank představuje hodnotu důvěryhodnosti tak, že závisí na počtu stránek, které jsou současně hodnoceny jedním vzorcem a na danou stránku odkazují. Čili každá webová stránka předává část své "hodnověrnosti" stránkám, na které odkazuje. Sama stránka tímto "předáváním" o svou hodnověrnost nepřichází, jakoby jí pouze přeposílá dále.

Čím více odkazů na stránce existuje, tím menší je předávaná hodnota PR každého z nich. V praxi to znamená, že pokud na naši stránku bude odkazovat stránka s PR5, která celkem obsahuje 10 odkazů, včetně odkazu na naši stránku, bude to pro nás přínosnější, než když na naši stránku bude odkazovat stránka s PR7, která ale bude obsahovat celkem 100 odkazů. Tento systém je v jednoduchosti znázorněn na obrázku 4.

Na obrázku 4 vidíme pět stránek označených velkými písmeny A-D a jejich hodnocení PR v kolečku. Hodnota uváděná u šipek, je hodnota PR odkazu na další stránku. Na stránku A směřují dva odkazy, jeden má hodnotu 2 a druhý hodnotu 1. Konečné ohodnocení stránky A je tedy 3. Stránka A také odkazuje na tři další stránky a každý z těchto odkazů má hodnotu 1. Na stránku D směřují také dva odkazy, ovšem oba mají jen hodnotu 1, a proto má tato stránka hodnotu pouze 2.

<sup>8</sup>Kubíček, s. 44.



Obrázek 4: Předávání PageRanku mezi stránkami

Musíme si však uvědomit, že PageRank je relativní hodnota, která může postupem času stoupat i klesat a je důležitým, nikoli však jediným faktorem, ovlivňujícím výslednou pozici webové stránky na stránce výsledků vyhledávání.

### 3.5.2 S-rank - hodnocení dle Seznamu

S-rank webové stránky je podobně jako u PageRanku od Google veličina vyjadřující důležitost stránky v prostředí českého internetu. Hodnocení je založeno na síti odkazů. Algoritmus počítá jednak s odkazy, které ukazují na stránku, ale zohledňuje také odkazy umístěné na stránce, které odkazují jinam. Přesné výpočty S-ranku však nejsou veřejné. Principy hodnocení stránek si vyhledávače pečlivě střeží jako své tajemství. Přesto je známo, že S-rank počítá nelineární kombinaci různých veličin, kde převažuje okolí a okolní vlivy působící na stránky (například zpětné odkazy) oproti tzv. on-page faktorům, což jsou konkrétní prvky na webové stránce (například titulek stránky, meta description, nadpisy atd.).<sup>9</sup>

<sup>9</sup>Kubíček, s. 45.

### 3.6 Sestavení výsledků vyhledávání

Stránku s výsledky vyhledávání, nebo také SERP (Search Engine Result Page) si každý vyhledávač tvoří automaticky sám a jiným způsobem. Kromě přirozeného vyhledávání, jsou zde zařazeny i placené reklamy, které jsou odlišeny, či zvýrazněny.<sup>10</sup> Hlavními prvky jsou:

- Titulek
- URL adresa
- Úryvek textu
- Náhled stránky

Titulek se shoduje s titulkem stránky, který jeho tvůrce ve zdrojovém kódu vepíše do značek `<title>` a `</title>`. Je to prvek, který je v SERPu nejvýraznější a nejvíce tak upoutá pozornost. Po jeho přečtení by měl být vyhledávající přesvědčen, že se jedná o stránku, kterou hledá.

URL adresa, nebo také webová adresa se většinou nachází pod titulkem. Jedná se o odkaz na konkrétní stránku, kde se nalézá odpověď na dotaz vyhledávajícího. Může se také stát, že odkaz na stránku je přeměrován přes databázi vyhledávače, který si tímto mimo jiné může pomoci prokliků kontrolovat relevantnost svých výsledků vyhledávání. Pokud se nějaká stránka nalézá na prvních příčkách ve vyhledávání, ale i přes to se jí většina návštěvníků vyhýbá, může této stránce vyhledávač snížit hodnocení a umístit ji na horší pozici.

Úryvek (anglicky snippet) je část textu ze stránky, který je podle vyhledávače nejvhodnější pro hledanou frázi nebo klíčové slovo. Každý vyhledávač tuto část textu sestavuje svým vlastním způsobem. Tři obecné možnosti sestavení snippetu, které mohou být i kombinovány<sup>11</sup>:

- Sestavení přímo z obsahu zaindexované stránky. Některé vyhledávače zobrazí prvních pár slov ze začátku stránky, jiné používají sofistikovanější metodu. Analyzují obsah stránky ze své databáze a vyberou části vět nacházejících se kolem klíčových slov nalezených na stránce. Často také v kombinaci s textem ve značce description.
- Úryvek je utvořen obsahem meta značky description (popis), která se nachází ve zdrojovém kódu stránky a vyplňuje jí samotný autor. Tento způsob se však může stát předmětem spamu a místo smysluplného popisu stránky mohou značky obsahovat pouze výčet nerelevantních klíčových slov, aby přilákaly návštěvníky na podvrženou, či virem infikovanou stránku. Z tohoto důvodu těmto značkám vyhledávače nepřikládají tak velkou důležitost.
- Dalším typem snippetu může být popis z katalogu. Například fulltextový vyhledávač Seznam, má přístup k vlastní databázi katalogu, odkud vyhledává popisky.

<sup>10</sup>Kubíček a Linhart, s. 21.

<sup>11</sup>Kubíček a Linhart, s. 22.

Může se stát, že pokud je tento popis relevantnější, než jakýkoliv jiný text na stránce, použije se jako úryvek na stránku s výsledky vyhledávání.


Náhled stránky je ukládán a následně zobrazován automaticky vyhledávačem. O jeho uložení a průběžnou aktualizaci se stará speciální robot (crawler), který stránku navštívuje společně s robotem "stahovačem". Typy náhledů mohou být různé, například Seznam ukládá z levého horního rohu oříznutou část stránky, kterou zmenší do velikosti 100 x 75 pixelů. Takto upravený náhled je pak zobrazován na stránce vyhledávání u každého nalezeného výsledku. Naopak Google pouze po najetí kurzorem myši na výsledek vyhledávání, zobrazí tlačítko náhledu, po jehož interakci se zobrazí náhled celé stránky s vyznačeným textem ze snippetu, nebo jiným relevantním obsahem.

Na obrázku 5 můžeme vidět sestavenou stránku s výsledky vyhledávání ve vyhledávači Seznam na klíčové slovo "monitor". Kromě čtyř výše popsanych prvků u přirozeného vyhledávání, na obrázku vidíme placenou reklamu. Indikuje ji žlutě zvýrazněný obdélník na začátku stránky. Další reklama je umístěna napravo od přirozeného vyhledávání. Google i Bing sestavují své stránky s vyhledáváním velice podobně, ovšem Bing nepodporuje placené kampaně, nenajdete tudíž zde žádné placené odkazy.

Internet [Firmy](#) [Mapy](#) [Slovník](#) [Zboží](#) [Obrázky](#) [Videa](#) [Encyklopedie](#)


**SEZNAM.CZ** monitor  [Pokročilé hledání »](#)

► [Všechny](#) [Česky](#) [Cizojazyčné](#)




**LCD monitor | Alza.cz**  
Rozjedte to s novým LCD monitorem! Filmy, hry i grafika budou zážitek.  
[Alza.cz/Lcd-Monitor](#)


Reklama Sklik




**Porovnání cen e-shopů, srovnání cen, monitor.cz**  
Monitor.cz je srovnávač cen internetových obchodů - eshopů. Porovnává ceny nabízených produktů z jednotlivých eshopů. Srovnat ceny můžete z nabídky více jak ...  
[www.monitor.cz/](#) - Praha 3 [Zobrazit na mapě](#)




**Našli jsme monitor » Zboží.cz**  
- Philips 166V3LSB (166V3LSB/00) 5 výrobků  
- BENQ G2025HDA černá barva 20 výrobků  
- BENQ G925HDA 4 výrobky  
[Další výrobky »](#)



**Monitory - Heureka.cz**  
Heureka.cz vám poradí jak vybírat **Monitory**. Vybírejte si **Monitory** podle parametrů a srovnávejte ceny z internetových obchodů na Heurece. ... LCD **monitor** (1194)  
[monitor.heureka.cz/](#)



**LCD monitor, CZC.cz**  
LCD **monitor**, LED, 3D, široká nabídka úhlopíček. Nakupujte u odborníků na počítače a elektroniku. ... Samsung SyncMaster S24B300BLZ - LED **monitor** 24"  
[www.czc.cz/lcd-monitor/produkty](#)



**Monitor (obrazovka) - Wikipedia**  
[Základní parametry monitorů](#) [Technologie zobrazení](#) [Některé významní výrobci](#)  
Většina výrobců totiž udává velikost úhlopíčky včetně plochy, kterou skryje plastový rám monitoru, a jelikož je tato plocha hlavně u CRT monitorů velká i ...  
[cs.wikipedia.org/wiki/Monitor\\_\(obrazovka\)](#)

Reklama Sklik:

[Chcete nový LCD monitor?](#)  
Pořídte ho za super cenu! **Monitory** se slevou. Možné i na splátky.  
[Gigacomputer.cz/Monitory](#)  
Masarykovo náměstí 3090/15, Ostrava

**Monitor z Aukra**  
**Monitor** pro Vás.  
Rychlý prodej, pohodlný nákup.  
[Aukro.cz](#)

**Monitory EIZO**  
Špičkové **monitory**, které maximálně šetří zrak. Různé možnosti využití.  
[Eizo.cz](#)  
K Závěrci 2751/32, Praha

[Srovnávejte než nakoupíte](#)  
Vybírejte podle parametrů a recenzí  
Nakupte za nízké ceny na Heurece  
[Heureka.cz](#)

**Monitory**  
Pořídte si nový **monitor**.  
Skvělé ceny a rychlé dodání.  
[Kasa.cz/Monitory](#)

[Tady může být i vaše reklama »](#)

Obrázek 5: Ukázka výsledků vyhledávání - Seznam

### 3.7 Klíčová slova

Klíčová slova (anglicky keyword) nebo fráze stručně popisují obsah. Například mohou být využity jako metadata (popisují jiná data, poskytují informace o obsahu určité položky) popisující obrázky, textové dokumenty nebo webové stránky. Dokumenty takto označené relevantními klíčovými slovy mohou být později vyhledány mnohem jednodušeji.<sup>12</sup>

#### 3.7.1 Relevance

Relevantní klíčové slovo je takové, které co nejlépe vystihuje obsah, nebo záměr daného dokumentu. Díky dobře cíleným a relevantním klíčovým slovům mohou vyhledávající na internetu efektivně nalézt požadovaný obsah.<sup>13</sup> Pro konkrétní představu uvedu následující příklad:

„Provozujete malý hotýlek někde v Dolní Lomné. Samostatné slovo hotel je samozřejmě vysoce konkurenční. Optimalizovat stránky na toto slovo bude velmi náročné. Finančně i časově. Navíc je nepravděpodobné, že by turista hledajícího ubytování konkrétně v Beskydech, případně přímo v Dolní Lomné, hledal hotel dotazem čistě jen za použití slova hotel. Spojení jako:

- hotel Dolní Lomná
- hotel v Dolní Lomné
- hotel Dolní Lomná Beskydy
- hotel Třinecko Beskydy
- hotel Třinecko

budou použita s mnohem větší pravděpodobností a výsledky budou také přesnější. Návštěvník nalezne na tato slovní spojení relevantnější a kvalitnější výsledky.“<sup>14</sup>

#### 3.7.2 Long tail

Jako dlouhý ocas (anglicky long tail) jsou označována klíčová slova, nebo delší fráze, které nejsou ve vyhledávacích zadávány tak často jako obecná slova, avšak celkový součet jejich vyhledávání převyšuje hledanost slov obecných.<sup>15</sup> Jako přirovnání nám poslouží kometa, kde jádro komety tvoří obecná slova jako:

- Fotoaparát
- Hotel

<sup>12</sup>TECHTERMS.COM. *TechTerms.com* [online]. 2013 [cit. 2013-04-09]. Dostupné z: <http://www.techterms.com/>

<sup>13</sup>Kubíček, s. 93.

<sup>14</sup>Kubíček, s. 98.

<sup>15</sup>Kubíček a Linhart, s. 31.

- Hračky

a za ním táhnoucí se dlouhý ohon je tvořen méně zadávanými slovy, nebo frázemi. Například:

- Fotoaparát s automatickým zaostřováním
- Hotel v Praze
- Hračky pro malé děti

Je důležité si uvědomit, že není nejpodstatnější být na nejvyšších pozicích ve vyhledávání pro pár hlavních klíčových slov. Jde o to získat dobré pozice u velkého počtu příbuzných slov, slovních spojení dobře charakterizujících obsah stránky a podrobných dotazů.

## 4 Existující aplikace pro sledování pozic ve vyhledávači

### 4.1 Přehled

Jak v České republice, tak i v zahraničí existuje spousta těchto nástrojů lišící se především funkcionalitou, komplexností řešení a také omezením neplacených verzí. Kromě námi požadovaného sledování pozic ve vyhledávačích také nabízejí různé analýzy, jako jsou: analýza zdrojového kódu a klíčových slov, zjištění síly a kvality webu, kontrola zpětných odkazů. Firmy vyvíjející tyto projekty většinou také nabízejí svá samozřejmě placená školení. Pro účely této práce si nyní podrobně popíšeme 4 aplikace, z čehož budou dvě tuzemské a dvě zahraniční. Zaměříme se hlavně na oblast sledování pozic ve vyhledávačích v neplacených verzích aplikací.

### 4.2 SEO PowerSuite

Jedná se o balík čtyř nástrojů, který spravuje firma Link-Assistant.com. Podle vývojářů jde o "vše v jednom" software, zabývající se optimalizací pro vyhledávače. Kompletní balíček lze testovat ve zkušební verzi, nebo koupit za \$249 ve verzi "Professional", nebo za \$599 verzi "Enterprise", která je vhodná spíše pro větší společnosti. Všechny verze se od sebe značně liší, například v maximálním počtu sledovaných konkurenčních webových stránek. Ve zkušební verzi není možné ukládat získané výsledky a kontrolovat tak historii vývoje. Naopak nejdražší "Enterprise" verze nabízí navíc generování různých druhů reportů. Ve stejných verzích lze také zakoupit jednotlivé nástroje samostatně a to za \$99,75 nebo \$249,75.<sup>16</sup>

#### 4.2.1 Rank Tracker

První nástroj se zabývá sledováním pozic ve vyhledávačích, a analýzou klíčových slov. Po nainstalování a spuštění zkušební verze jsme vyzváni k zadání URL adresy, na kterou se chceme zaměřit. Následuje výběr vyhledávačů, ve kterých chceme nechat náš web otestovat. Množství podporovaných vyhledávačů je nepřehledné a dělí se podle států. Dále je zde možnost zadat klíčová slova, jejichž počet není ani v neplacené verzi nějak omezen. Na konec nastává skenování, které trvá řádově v jednotkách až desítkách minut v závislosti na počtu zadaných klíčových slov. Nyní se dostaneme do hlavního přehledu projektu, kde si můžeme přidat další klíčová slova, zadat API klíče vyhledávačů anebo přidat maximálně jeden konkurenční web. Výsledkem je report v podobě grafu a tabulky, kde pro každé klíčové slovo a každý vyhledávač je zobrazena pozice. Veliké omezení této neplacené verze však tkví v nemožnosti uložení dosažených výsledků, což naprosto znemožňuje dlouhodobé sledování vývoje pozic.

<sup>16</sup>LINK-ASSISTANT.COM. *SEO PowerSuite* [online]. 2013 [cit. 2013-03-26]. Dostupné z: <http://www.link-assistant.com/>



#### 4.2.2 WebSite Auditor

Tento nástroj má na starost dotvářet, nebo upravovat webové stránky tak, aby byly pro vyhledávače co nejpřijatelnější a tím pádem získaly dobré hodnocení a vysoké příčky ve vyhledávání. Nástroj skenuje všechny webové stránky, které se nacházejí pod doménou. Odstraní všechny problémy bránící dobrým výkonům stránek a pomůže Vám je optimalizovat pro každý vyhledávač a každé klíčové slovo. Hlavní problémy, kterými se WebSite Auditor zabývá, jsou:

- Nefunkční odkazy
- Chyby v HTML kódu
- Duplicitní obsah
- Neindexované stránky
- Chybějící, nebo příliš dlouhé titulky a popisy

#### 4.2.3 SEO SpyGlass

Funkcí SEO SpyGlass je analýza konkurenčních webových stránek, které jsou hodnoceny velice dobře. To znamená, že jejich strategie budování zpětných odkazů je dobrá a můžete tak zjistit, kde se nachází spousta kvalitních a dobře hodnocených zpětných odkazů.

#### 4.2.4 LinkAssistant

Tento nástroj se zaměřuje na tvorbu kampaní pro výměnu nebo nákup zpětných odkazů. Automaticky vyhledává vhodné partnerské webové stránky a snaží se získat kontakt na majitele. Generuje vlastní seznam zpětných odkazů a zároveň kontroluje, zda jsou všechny odkazy funkční. Zjišťuje také, zda se na již smluvených partnerských webových stránkách nachází funkční zpětný odkaz na Vaše stránky bez použití nepovolených SEO praktik.

### 4.3 Web CEO

Druhým nástrojem je projekt Web CEO firmy Web CEO Ltd. V nabídce je desktopová i webová aplikace. Desktopová stojí jednorázově \$499 + volitelné tříměsíční poplatky například za aktualizace znalostní báze vyhledávačů. Existuje také 30-ti denní zkušební verze. Zatím co používání webové aplikace může stát od \$49 do \$299 měsíčně. Je k dispozici časově neomezená zkušební verze, ovšem s omezeními v počtu klíčových slov, konkurenčních webových stránek a vyhledávačů.<sup>17</sup>

Webová a desktopová aplikace jsou dva rozdílné produkty a koncepty:

<sup>17</sup>WEB CEO LTD. *Web CEO* [online]. 2013 [cit. 2013-03-26]. Dostupné z: <http://www.webceo.com/>

- U webové aplikace nejste vázáni na žádný software instalovaný na počítači ani na místě odkud chcete ke svým projektům přistupovat. Kdekoliv, kde máte přístup k internetu, můžete také přistupovat ke svým datům.
- Procesy webové aplikace běží na serverech firmy Web CEO Ltd. nemůže se tedy stát, že ztratíte data vinou zničení počítače. Znalostní báze vyhledávačů bude stále aktuální.
- Ačkoliv se sada nástrojů ve webové a desktopové aplikaci mírně liší, základní nástroje jsou obsaženy v obou produktech.
  - the Site Rank Checker (kontrola pozic ve vyhledávačích)
  - the Visitors activity analysis (analýza aktivit návštěvníků)
  - the Site Auditor (kontrola validity webových stránek)
- Webová aplikace je určena jak pro začátečníky, tak pro profesionály, zatím co desktopová aplikace je určena výhradně profesionálům.

Nástroje webové aplikace:

- Buzz Tracker (správa zpětných odkazů)
- To Do List (úkoly k řešení)

Nástroje desktopové aplikace:

- HTML Editor (editace HTML kódu stránek)
- Website Uptime Monitoring (neustálá kontrola dostupnosti stránek)

Po registraci a následném přihlášení k online aplikaci budeme vyzváni k vytvoření nového projektu, kde se zároveň vyplňuje URL adresa naší webové stránky. Dále pokračuje nabídka nastavení, kde je možno zadat potřebná klíčová slova, vyhledávače a URL adresy konkurenčních webů. I zde je na výběr velké množství vyhledávačů seřazených a rozdělených dle států. Možnost zadat až 20 konkurenčních webů k následnému porovnání, je poměrně dobrá. Na druhou stranu naprosto neuspokojivé je omezení na zadání maximálně pěti klíčových slov. Ve většině případů bude tento počet velice omezující. V neplacené verzi je také zakázán přístup k nastavení časového plánu pro skenování pozic. Uživatel je tak odkázán sám na sebe a skenování musí spouštět manuálně. Konečná prezentace nashromážděných dat je zobrazena pomocí několika oddělených tabulek a grafů:

- Tabulka vlastních pozic - seřazena podle vyhledávačů a v nich vyhledaných klíčových slovech. Obsahuje pozici posledního skenování doplněnou o změnu oproti minulému včetně malého orientačního grafu vývoje.
- Tabulka pozic konkurentů - obdoba předchozí tabulky, jen zobrazuje informace o konkurenčních webových stránkách

- Tabulka pozic konkurentů seřazená dle klíčových slov - další obdobná tabulka ovšem třízená s ohledem na klíčová slova
- Graf vlastních pozic - vykresluje graf závislosti pozic ve vyhledávači na časové ose pro každý vyhledávač zvlášť. Další údaje do grafu jsou zaznamenány při každém dalším ručním skenování
- Graf pozic konkurentů - zde je opět zobrazen graf závislosti pozic ve vyhledávači na času oproti předchozímu jsou zde zobrazeny křivky všech konkurenčních webů, avšak hodnoty jsou počítány jako průměr všech klíčových slov ve všech vyhledávacích

#### 4.4 Seo Servis

První českou aplikací je Seo Servis firmy Webový servis Company s.r.o. je ryze webovou aplikací a její ovládání je intuitivní. Modul pro zjišťování pozic ve vyhledávači je poněkud strohý a nabízí pouze tři možnosti vkládání parametrů. Konkrétně jde o Vaší doménu, neboli URL adresu webové stránky, klíčové slovo a jazykovou specifikaci. Po zadání těchto parametrů automaticky vyhledá pozice pro čtyři vyhledávače (Google, Seznam, Bing a Jyxo). Výstupem je jednoduchý report s informacemi o zvoleném jazyku, PageRanku a S-ranku. Dále zde najdeme tři jednoduché tabulky:

- Pozice ve vyhledávačích - obsahuje pouze čtyři řádky, každý pro jeden z vyhledávačů a příslušnou pozici s odkazem
- Zpětné odkazy a zaindexované stránky - obsahuje informace patrné z názvu tabulky a to opět pro každý vyhledávač zvlášť
- Pozice ve vyhledávačích v minulosti - informace o pozicích před jedním a třemi měsíci

Pro neregistrované uživatele může být nepříjemné ukládání pětadvacítiky posledních výsledků vyhledávání do databáze a jejich veřejné zobrazování na stránce. Toto však lze vypnout v nastavení po bezplatné registraci, díky které také získáte přístup k výsledkům Vašeho vyhledávání v minulosti.

#### 4.5 Seo nástroje

Poslední popsanou aplikací jsou Seo nástroje firmy SEO Linhart s.r.o. a je taktéž čistě webovou aplikací. Vstupní formulář obsahuje pouze dvě možnosti pro vložení parametrů a to pro frázi, neboli klíčové slovo a URL adresu. Výhodou je možnost zadání více klíčových slov nebo URL adres, každou na jeden řádek. Avšak omezení je nastaveno na počet dotazů, maximální počet je pět a jako dotaz je bráno každé klíčové slovo pro každou URL adresu, tedy pokud zadáme tři klíčová slova a dvě URL adresy, tedy celkem šest dotazů, aplikace nás upozorní na překročený počet a analýza se nespustí. Dalším velice nepříjemným omezením je tzv. captcha, což je ochrana formulářů sloužící pro

ověření, zda zadavatel je skutečně člověk a nikoliv škodlivý program. V podstatě jde o opsání zdeformovaného textu do kontrolního políčka, což je při potřebě většího počtu vyhledávání časově náročné. Toto omezení lze vyřešit bezplatnou registrací, po přihlášení již captcha zmizí, omezení v podobě maximálního počtu dotazů však zůstává i nadále.

## 5 Vlastní nástroj

### 5.1 Vize

Základem nástroje by měla být schopnost stahování a ukládání výsledků vyhledávání ze třech hlavních vyhledávačů Googlu, Seznamu a Bingu. Jedná se o výsledky přirozeného vyhledávání, kde hraje důležitou roli počet stažených výsledků, neboli počet pozic od první až po n-tou. Dále také placené kampaně, které se zobrazují zároveň s přirozeným vyhledáváním.

Takto uložená data by měla být přehlednými způsoby prezentována pomocí tabulek či grafů. Nesmíme opomenout možnosti selekce a filtrování prezentovaných dat, stejně tak obecné přehledy.

Aplikace musí být samostatná. Jelikož je plánována její distribuce jednotlivým zákazníkům, nesmí být k chodu aplikace zapotřebí žádných dalších podpůrných nástrojů, které by bylo nutno instalovat zároveň s aplikací.

Mělo by se jednat o kombinaci desktopové a webové aplikace. Způsob kombinace tkví v tom, že desktopová aplikace, je schopna plně pracovat sama o sobě s veškerou funkcionalitou. S webovým serverem bude pak komunikovat pouze dle potřeby. Půjde pouze o výměnu vstupních dat a získaných výsledků.

### 5.2 Specifikace požadavků

Jediný, kdo bude s aplikací pracovat, je uživatel. Ten musí mít možnost co nejvíce aplikaci ovládat a její funkčnost nastavit tak, aby přesně vyhovovala jeho požadavkům.

Vstupními daty (zadávány zákazníkem) jsou:

- Vlastní webové stránky
- Konkurenční webové stránky
- Klíčová slova
- Nastavení (počet stahovaných pozic, vyhledávače, placené kampaně, výsledky vyhledávání přes API)

Výstupní data (grafy a tabulky):

- Detailní pozice ve vyhledávači v závislosti na klíčovém slovu a datu
- Přehled více webových stránek v jednom vyhledávači
- Detailní přehled zobrazování placené reklamy pro určitou webovou stránku na zadané klíčové slovo
- Celkový pohled na výskyty placených reklam, závislosti webových stránek na datu
- Přehled klíčových slov v závislosti na vyhledávači

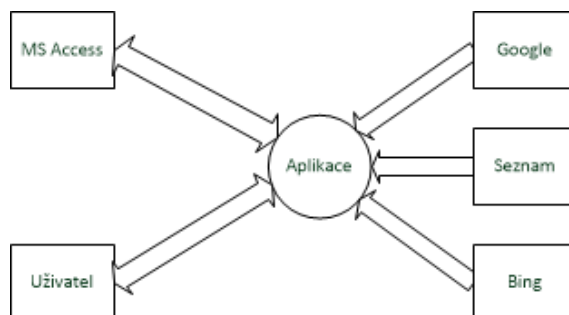
- Přehled jednotlivých URL adres webové stránky

Funkce aplikace:

- Stahování výsledků vyhledávání ze Seznamu (přirozené / placené)
- Stahování výsledků vyhledávání z Googlu (přirozené / placené, přes API / parsováním)
- Stahování výsledků vyhledávání z Bingu (přes API / parsováním)
- Roztřídění a uložení dat do databáze
- Příslušná prezentace dat
- Komunikace s webovou aplikací

### 5.3 Okolí aplikace

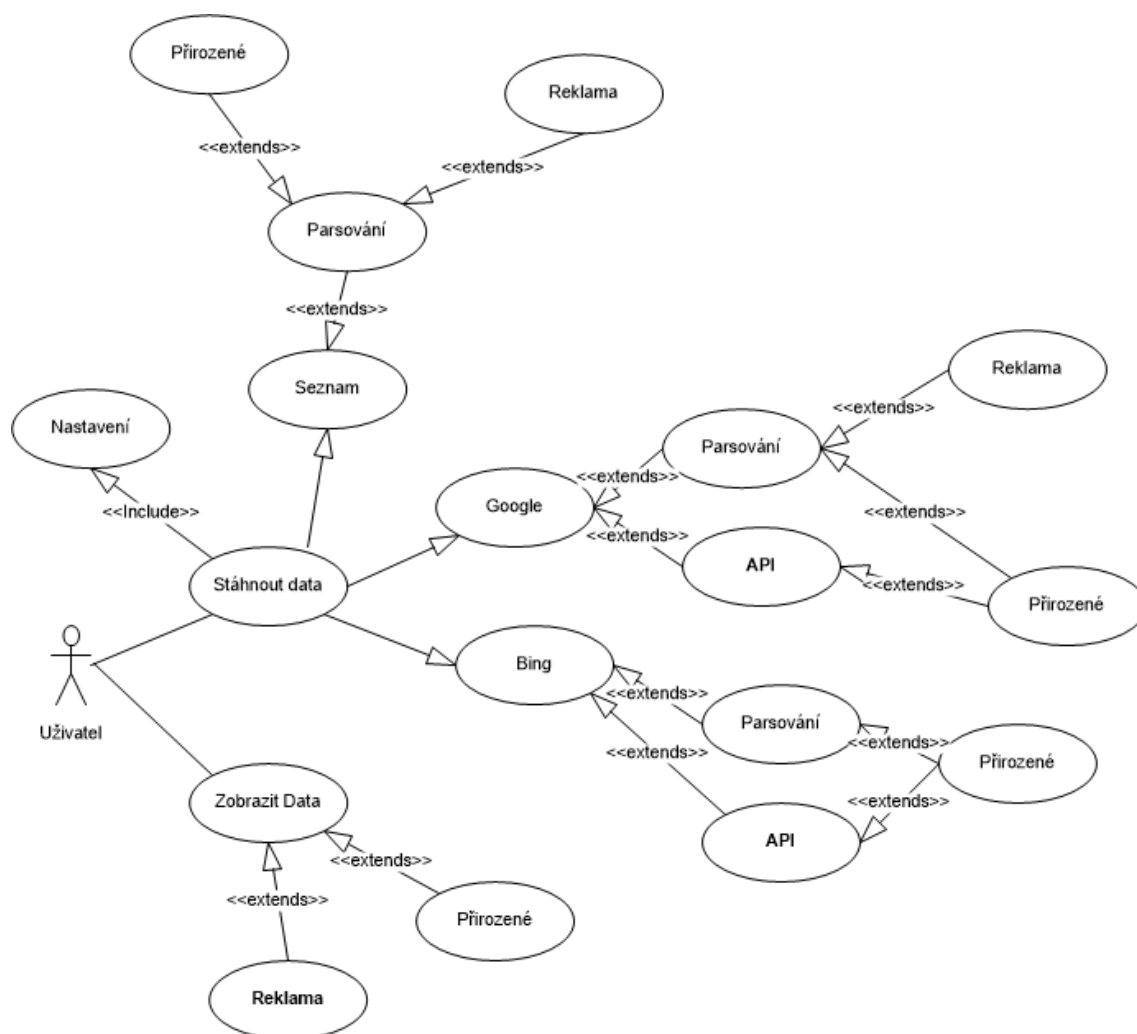
Na obrázku č. 6 je zobrazeno okolí aplikace. Bude s ní pracovat uživatel, který má možnost aplikaci ovládat a řídit, ale také získávat přehledy v podobě grafů a tabulek. Databáze MS Access slouží k uchovávání dat a je díky ní možné zobrazovat již zmíněné tabulky a grafy. Google, Seznam a Bing jsou vyhledávače, ze kterých aplikace získává data.



Obrázek 6: Okolí aplikace

### 5.4 Diagram případu užití

Diagram případů užití na obrázku č. 7 popisuje akce, které může uživatel provádět. Zobrazování dat je odděleno na zobrazování reklamních kampaní a přirozeného vyhledávání. Stahování dat, pak obsahuje nastavení a data je možno stahovat z Googlu, Seznamu a Bingu. Na diagramu je pak znázorněno, zda půjde o stahování přirozených výsledků či reklamních kampaní a zda se budou stahovat pomocí API, nebo parsováním.



Obrázek 7: Diagram případu užití

## 5.5 Stahování výsledků vyhledávání

### 5.5.1 Požadovaná data

Z výsledků vyhledávání nás zajímá hlavně pozice, URL adresa webové stránky na této konkrétní pozici a jeho popis. K tomuto musíme připojit klíčové slovo, pro které vyhledávání proběhlo, dále vyhledávač a také datum stažení.

U placených reklam budeme evidovat pouze URL adresu webové stránky, datum a na jaké klíčové slovo, ve kterém vyhledávači se placená reklama objevuje.

### 5.5.2 Parsování

Tato data lze získat více způsoby. Prvním je parsování zdrojového kódu stránky. Jde vlastně o stáhnutí celé stránky a následně pomocí HTML parseru vytržít pouze požadovaná data. Takto lze přistupovat jak k přirozeným výsledkům vyhledávání, tak i k placené kampani. Nevýhodou tohoto řešení je možnost dočasného zablokování ze strany vyhledávače. Pokud aplikace stahuje velké množství dat rychle za sebou, je vyhodnocena jako robot. Seznam pak chce ověřit, zda se jedná o uživatele použitím captcha, Google rovnou na několik hodin zablokuje přístup. Jediným řešením je pak čekání.

Druhým způsobem je přístup k výsledkům vyhledávání přes API. Nejprve si musíte u vyhledávače zařídit účet a získat specifický API klíč, díky kterému pak můžete přistupovat k datům. Nehrozí tak zablokování ze strany vyhledávače jako tomu bylo v předchozím případě, avšak přístup, bude omezen maximálním počtem dotazů. Získáme tak "čistá" data, bez rušivých elementů a nutnosti parsování celého zdrojového kódu webové stránky. Přicházíme však o možnost získání výsledků placených kampaní. Ne každý vyhledávač tento přístup podporuje. Z naší trojice nemá API pro vyhledávání Seznam.

V aplikaci jsou implementovány oba způsoby. Uživatel si tak může vybrat, zda si sám vytvoří účet a získá vlastní API klíč, nebo bude stahovat data pomalejším parsováním zdrojového kódu webové stránky. Následující ukázka popisuje způsob vyhledání a získání dat přirozeného vyhledávání Googlu pomocí parseru HTML Agility Pack.<sup>18</sup>

```
private int DC = DatabaseAcces.DownloadCount;
public List<Item> Download(string keyword)
{
    int pozition = 1;
    int repeteCycle = 0;
    bool urlOrDesc = true;
    List<Item> polozky = new List<Item>();
    Item polozka = new Item();
    HtmlAgilityPack.HtmlDocument htmlDoc = new HtmlAgilityPack.HtmlDocument();
    String urlTemplate = @"https://www.google.cz/search?q={0}&start={1}&oe=utf8";
    for (int i = 0; i < DC; i += 10)
    {
        try
        {
            Uri searchUri = new Uri(String.Format(urlTemplate, keyword, i));
            String page = new WebClient() { Encoding = Encoding.UTF8 }.DownloadString(searchUri);
            htmlDoc.LoadHtml(page);
            var divs = htmlDoc.DocumentNode.SelectNodes("//li[@class='g']/div[@class='s']/div[@class='kv']/cite_|_//li [ @class='g']/div [ @class='s']/span[@class='st']");
            foreach (var item in divs)
            {
                if (urlOrDesc)
                {
                    polozka.URL = item.InnerText;
```

<sup>18</sup>MICROSOFT. *Html Agility Pack* [online]. 2013 [cit. 2013-04-26]. Dostupné z: <http://htmlagilitypack.codeplex.com/>



```

        urlOrDesc = false;
    }
    else
    {
        polozka.Description = item.InnerText;
        polozka.Position = pozition++;
        polozky.Add(polozka);
        urlOrDesc = true;
    }
}
}
catch (Exception exc)
{
    i -= 10;
    repeteCycle++;
    if (repeteCycle > 10)
    {
        repeteCycle = 0;
        throw exc;
    }
}
}
return polozky;
}

```

#### Výpis 1: Stahování a parsování dat

Obdobně, dle této ukázky, je řešeno i stahování dat z ostatních vyhledávačů a to jak placených kampaní, tak přirozeného vyhledávání. Odlišným způsobem jsou řešená stahování přes API. Liší se však pouze doplněním ověřovacích údajů (API klíč, uživatelské jméno, heslo) a k vyřízení jednotlivých dat je použita třída XmlNode.<sup>19</sup>

Ve zdrojovém kódu jsou nejprve inicializovány potřebné proměnné a vytvořeny instance tříd. Proměnná "urlTemplate" je šablona URL adresy vedoucí na stránku s výsledky vyhledávání, ve které jsou předpřipravena místa pro zadání klíčového slova a pozice, od které jsou výsledky zobrazeny. Následuje cyklus s předem daným počtem opakování, který je iterován po deseti. V každém cyklu se provede stažení zdrojového kódu stránky, na které se mimo jiné nachází i deset výsledků vyhledávání. Ty je potřeba oddělit od ostatního obsahu. Do proměnné "divs" jsou pomocí jazyka XPath a metody "SelectNodes" obsažené v Html Agility Packu, uloženy obsahy jednotlivých elementů ze zdrojového kódu stránky. Do pole jsou střídavě uloženy URL adresa a popis webové stránky. Následně se cyklem foreach projdou všechny prvky pole a do proměnné "polozka" se postupně uloží URL adresa, pozice a popis webové stránky. Proměnná "polozka" je pouze pomocná struktura pro ucelení těchto tří atributů, která je následně přidána do seznamu položek. Ten je po úspěšném stáhnutí předán k dalšímu zpracování.

<sup>19</sup>MICROSOFT. *MSDN Library* [online]. 2013 [cit. 2013-04-30]. Dostupné z: <http://msdn.microsoft.com/en-us/library>

### 5.5.3 Databáze

Proto, aby na uživatelskou stanici nebylo nutné instalovat SQL server, byl jako SŘBD zvolen Microsoft Access. Stažené výsledky jsou tedy ukládány do předpřipraveného souboru s tabulkami, který je součástí aplikace. Pro hlavní funkci aplikace je nezbytná tabulka "storedResults", kde jsou uchovávána veškerá potřebná data. Po stažení veškerých výsledků do tabulky "stazeno" jsou z ní vybrána pouze požadovaná data, které se zkopírují do tabulky "storedResults". Ostatní nepotřebná data jsou smazána. Tímto způsobem funguje i ukládání výsledků placených kampaní. Dále jsou v soboru předpřipraveny tabulky pro seznam klíčových slov, vlastních a konkurenčních domén a jejich jednotlivých URL adres.

V datovém slovníku jsou popsány struktury uložených tabulek v databázi včetně atributů a jejich vlastností (název, datový typ, primární klíč, cizí klíč, null). Kompletní datový slovník se nachází v příloze a jeho ukázkou je tabulka č.3 .

storedResults				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Keyword	Text	NE	NE	NE
Datum	Datum a čas	NE	NE	NE
Searcher	Text	NE	NE	NE
Position	Číslo	NE	NE	NE
URL	Text	NE	NE	NE
Description	Memo	NE	NE	ANO

Tabulka 3: Ukázka datového slovníku, tabulka storedResults

## 5.6 Zobrazení výsledků vyhledávání

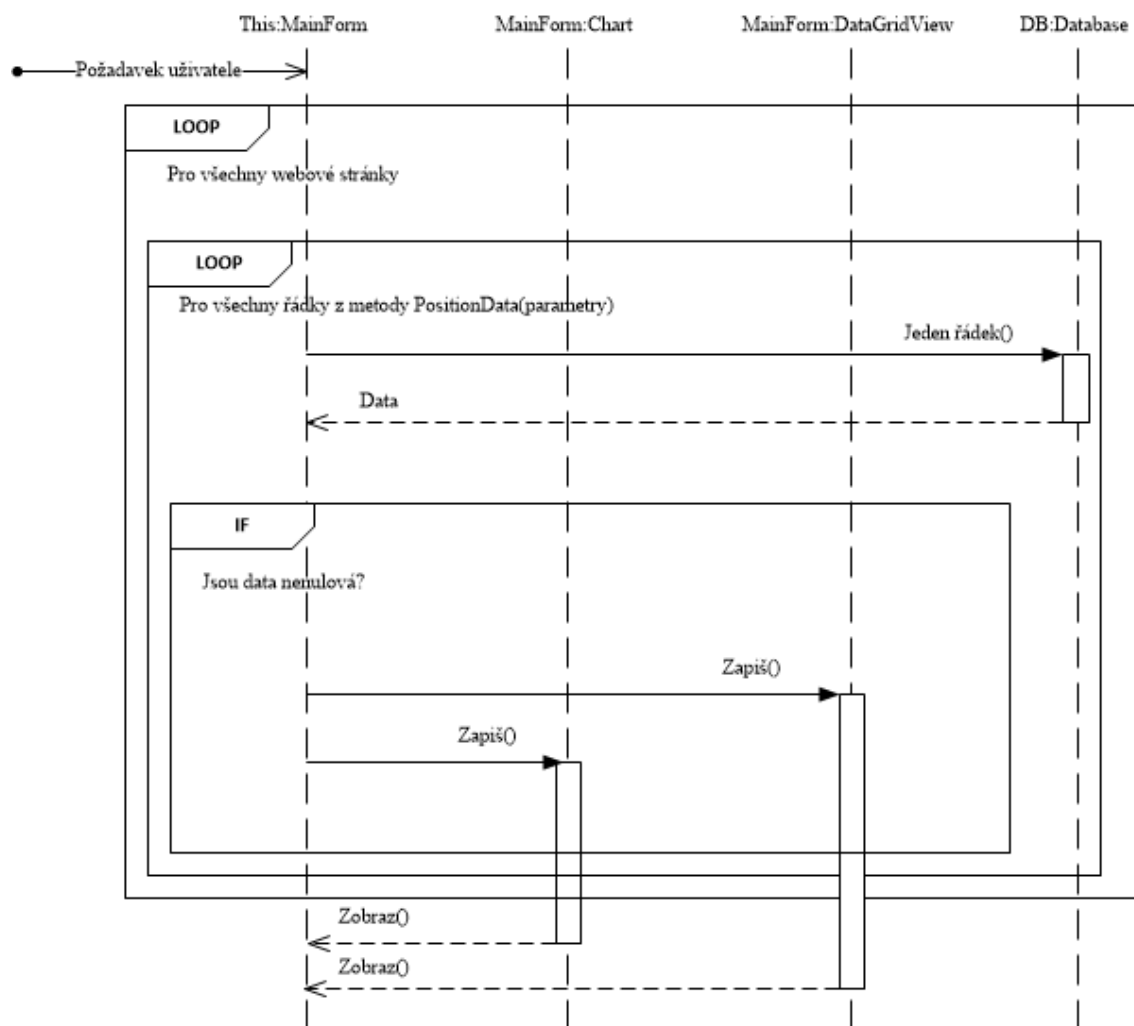
Pro prezentování již stažených dat jsou použity komponenty Chart a DataGridView. Uživatel nejprve provede potřebná nastavení a určí, co chce zobrazovat. Aplikace poté z databáze získá potřebná data a zobrazí je.

Základním stavebním kamenem pro každé zobrazení je tento SQL dotaz:

```
SELECT * FROM storedResults
WHERE (url LIKE @url)
AND (keyword LIKE @keyword)
AND (searcher LIKE @searcher)
AND (datum BETWEEN @DateOd AND @DateDo)
ORDER BY datum, [position] DESC
```

Výpis 2: SQL dotaz pro výběr dat

Vybírá všechny sloupce z tabulky StoredResult podle několika kritérií, které určují parametry (slova před kterými je zavináč, např. @url). Vyhledává se podle URL adresy webové stránky, klíčového slova, vyhledávače a časového rozmezí. Skládáním takto získaných dat lze vytvořit různé přehledy. Například způsob sestavení přehledu všech webových stránek pro jedno klíčové slovo v daném vyhledávači je znázorněn sekvenčním diagramem na obrázku č. 8.



Obrázek 8: Sekvenční diagram - sestavení přehledu

## 5.7 Komunikace s webovou aplikací

Tato desktopová aplikace je schopna fungovat sama o sobě. Její součástí však je kooperace s webovou aplikací. Je jakýmsi doplňujícím modulem. Pokud tedy uživatel pracuje s webovou aplikací, může si nechat vygenerovat klíč, po jehož předání dojde k synchronizaci dat mezi těmito aplikacemi. Vize je tedy taková, že do databáze desktopové aplikace jsou uložena vstupní data (klíčová slova, vlastní a konkurenční webové stránky) získaná z webové aplikace, která jsou následně zpracována a výsledky jsou odeslány zpět webové aplikaci. Zde jsou tato data shromážděna a připravena k dalším analýzám. Vzhledem k fázi vývoje této externí webové aplikace, která není předmětem této práce, je prozatím implementováno pouze získání dat.

Komunikace je řešena pomocí formátu JSON(JavaScript Object Notation), neboli zápis

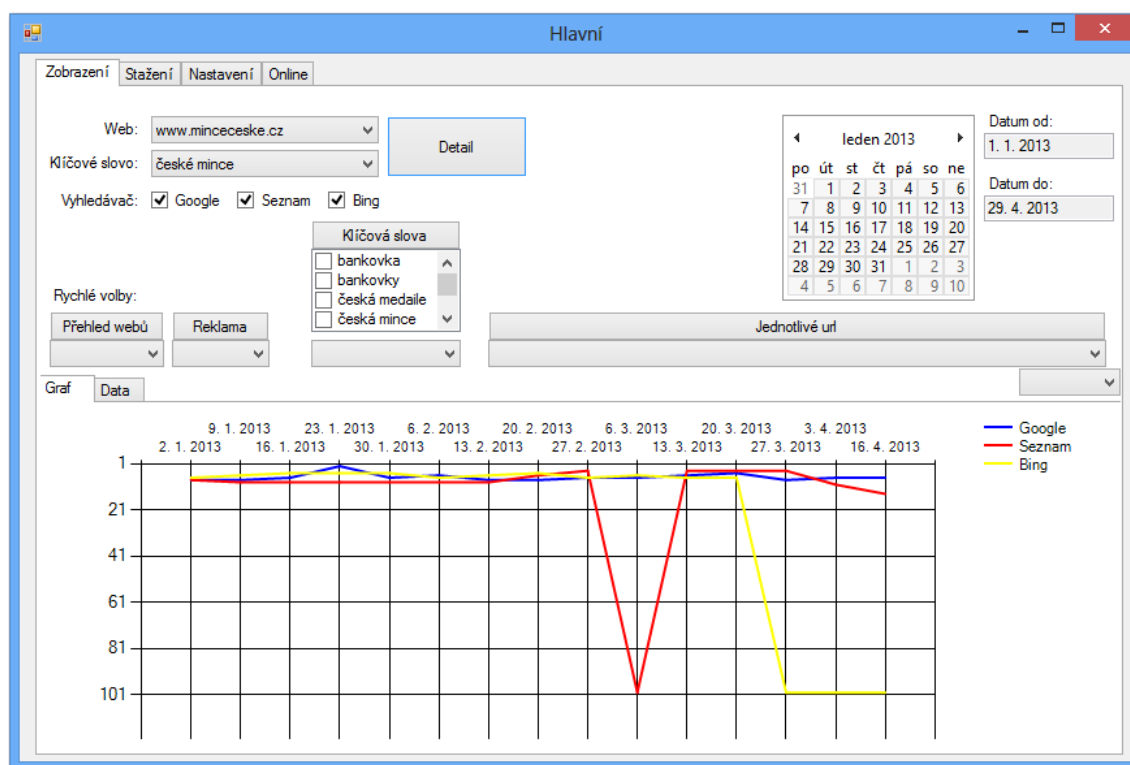
objektů pomocí JavaScriptu. Pro práci s tímto formátem byla použita knihovna Json.NET.<sup>20</sup>

## 5.8 Výsledná aplikace

K výsledné aplikaci musí být pro správnou funkci přiložen soubor s uloženým nastavením "apis.xml" a databázový soubor "DownloadedData.accdb". Dále také výše zmíněné knihovny Json.NET a Html Agility Pack. Samotná aplikace poté obsahuje čtyři záložky:

- Zobrazení

Zde jsou umístěny ovládací prvky pro výběr zobrazovaných dat včetně dynamicky generovaného grafu a tabulky. Konkrétní ukázka s testovacími daty lze vidět na obrázku č. 9.



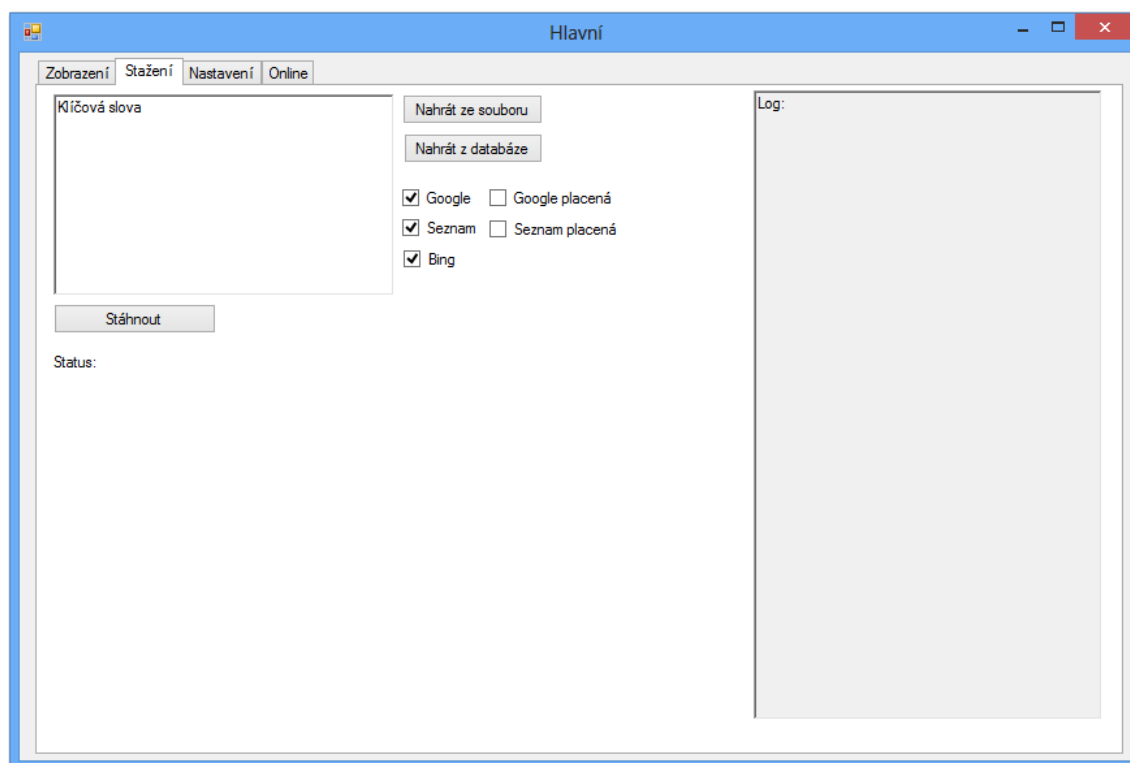
Obrázek 9: Ukázka aplikace - záložka Zobrazení

- Stažení

Tato záložka slouží pro stahování výsledků vyhledávání, které jsou po stažení a zpracování viditelné na předchozí záložce. Uživatel má možnost zadat klíčová

<sup>20</sup>JAMES NEWTON-KING. *Json.NET* [online]. 2013 [cit. 2013-04-26]. Dostupné z: <http://james.newtonking.com/projects/json-net.aspx>

slova, pro která chce získat výsledky a zvolit ve kterých vyhledávačích hledat a jaký typ reklamy stahovat. Také zde lze pozorovat průběh stahování. Rozložení této stránky lze vidět na obrázku č. 10.



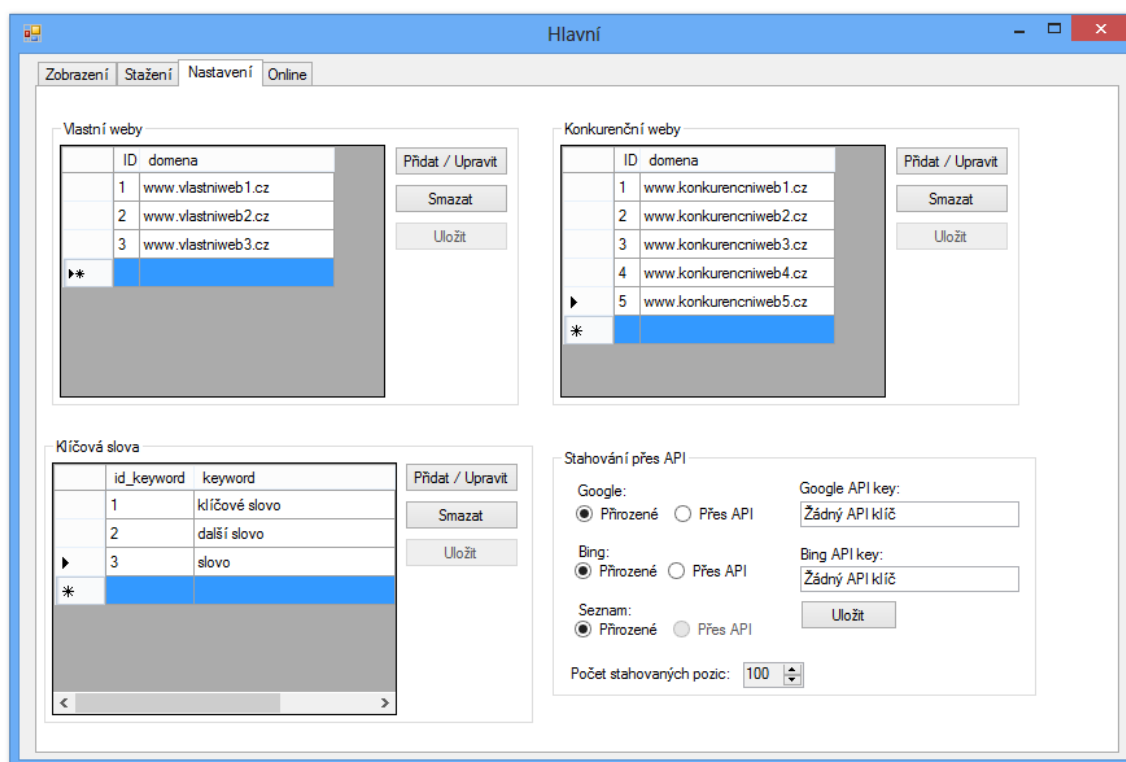
Obrázek 10: Ukázka aplikace - záložka Stažení

- Nastavení

Třetí záložka umožňuje spravovat vlastní a konkurenční webové stránky, klíčová slova, také nastavení počtu stahovaných pozic použití API klíčů. Záložka je zobrazena na obrázku č. 11.

- Online

Poslední záložka obsahuje pouze možnost zadání klíče pro komunikaci s externí webovou aplikací.



Obrázek 11: Ukázka aplikace - záložka nastavení

## 6 Srovnání s existujícími systémy

Abychom mohli zjišťovat, na jakých pozicích se umísťují naše, nebo konkurenční webové stránky, musíme je nejprve všechny zadat do systému. Dále pak také musíme určit, pro jaká klíčová slova chceme pozice zjišťovat a v jakých vyhledávačích. Zjištěné informace musí být uloženy a později dostupné k opětovnému prohlížení pomocí grafů či tabulek. Tyto možnosti však mohou být omezeny. Tabulka č. 4 uvádí přehled omezení, či maximálního počtu jednotlivých kategorií.

Srovnání takovýchto aplikací je velice těžké. Záleží totiž na individuálních požadavcích uživatele. Může se jednat o amatérského nadšence, kterému bude stačit, když se jednou za měsíc podívá, jak si jeho webová stránka ve vyhledávačích vede. Na druhou stranu může jít o profesionálního SEO specialistu, který spravuje spousty projektů svých klientů. V tomto případě už však lze uvažovat o zakoupení některého z produktů.

Počet klíčových slov, pro která bude běžný uživatel kontrolovat pozice ve výsledcích vyhledávání, bude řádově v desítkách. Bude mít jednu či dvě své webové stránky a několik konkurenčních. Z tohoto pohledu dobře vyhovují obě aplikace WEB Ceo, ale i vlastní aplikace. U webové aplikace WEB Ceo může být limitující počet klíčových slov. U aplikací Seo servis a Seo nástroje je také teoretická možnost sledování neomezeného počtu webových stránek avšak žalostným způsobem jednu po druhé. Navíc bez možnosti přehledného uložení historických dat. Jak je uvedeno v kapitole 3.4 podle statistik se nejčastěji používá tři až pět vyhledávačů. Uživatelů, kteří používají ostatní vyhledávače je zanedbatelně málo. Nemá tedy smysl kontrolovat pozice ve více než pěti vyhledávačích. Desktopová aplikace WEB Ceo jich má na výběr více než sedm set, což je určitě dobré pro distribuci v celém světě, ale uživatelé z jedné země stejně využijí pouze malý počet vyhledávačů, které jsou pro ně nejdůležitější. Z tohoto hlediska by tedy mohly být dostačující všechny srovnávané aplikace.

Důležitou součástí aplikace musí být možnost ukládání historických dat. Pouhé aktuální zjištění pozic jako je tomu u aplikací Rank Tracker a Seo nástroje bude ve většině případů nedostačující. Někomu vyhovují více tabulky a někomu zase graf. Ideální však je kombinace obou prvků. Toto podporuje pouze vlastní aplikace a webová aplikace WEB Ceo. Rank Tracker také podporuje obojí, ale bez možnosti uložení starších dat je graf zbytečný. Dalším nepříjemným faktem může být nutnost při každém spuštění aplikace znovu a znovu zadávat požadovaná klíčová slova a webové stránky.

Vlastní aplikace se svou funkcionalitou velice dobře přibližuje aplikacím WEB Ceo a Rank Trackeru. Její nespornou výhodou oproti neplacené verzi Rank Trackeru je možnost ukládání historických dat. Také není časově omezená na 30 dní jako zkušební verze desktopové aplikace WEB Ceo. Oproti webové aplikaci WEB Ceo zase nabízí vyšší maximální počet použitých klíčových slov a webových stránek. Nevýhodou vlastní aplikace může být horší grafické zpracování a méně přehledný proces stahování jednotlivých výsledků.

	WEB Ceo - Desktop	WEB Ceo - Web	Rank Tracker	Seo servis	Seo nástroje	Vlastní aplikace
Max. počet klíčových slov	292 MB <sup>a</sup> (sdíleno)	5	Neomezený	1	5/1	Neomezený
Max. počet vlastních web. stránek	292 MB (sdíleno)	2	1	1	1/5	Neomezený
Max. počet konkurenčních web. stránek	196 MB	40	1	Nerozlišuje	Nerozlišuje	Neomezený
Počet dostupných vyhledávačů	787	5	342	4	4	3
Historická data	Tabulky	Tabulky / grafy	Ne	Tabulky	Ne	Tabulky / grafy

Tabulka 4: Omezení aplikací

<sup>a</sup>Omezení není početní, ale maximum udává celková velikost uložených dat.



## 7 Závěr

Prvním cílem práce bylo seznámení s problematikou okolo SEO a existujících aplikací zabývajících se právě sledováním pozic ve výsledcích vyhledávání. K tomuto účelu bylo nutné prostudovat příslušnou literaturu, oficiální webové stránky vývojářů a také jednotlivé programy otestovat nad reálnými daty.

Dalším cílem této práce bylo vytvořit nástroj, který bude sledovat vývoj pozic ve výsledcích vyhledávání, se zaměřením jak na přirozené vyhledávání, tak i placené kampaně. Kromě vývoje pozic vlastních webových prezentací by měl být schopen sledovat také konkurenci. Nástroj je tedy vhodný nejen pro uživatele, kteří začínají optimalizací pro vyhledávače a chtějí sledovat vzestup své webové stránky na nejlepší pozice ve vyhledávání, ale také pro uživatele, kteří již jsou na dobrých příčkách ve vyhledávání, chtějí si tyto pozice udržet a navíc sledovat vývoj konkurence.

Výsledná aplikace je v tuto chvíli schopna získávat výsledky vyhledávání ze tří podílově nejvýznamnějších vyhledávačů v České republice (Google, Seznam, Bing) a to až do hloubky sté pozice. Do prvních tří stránek vyhledávání je také schopna stahovat placenou reklamu, která už se dále nemění. V aplikaci je také implementována možnost využít dva způsoby stahování výsledků vyhledávání a to přirozeně, nebo přes předpřipravená rozhraní vyhledávačů, která jsou k tomuto účelu stvořena. Takto stažené pozice jsou uloženy a pomocí různých typů přehledů následně prezentovány použitím grafů a tabulek.

Hlavní cíl práce tedy byl, dle mého názoru, splněn. Přestože aplikace by mohla být v budoucnu ještě rozšířena o další možnosti, které nemají ani tak vliv na funkcionalitu, ale spíše na usnadnění práce, či dobrou reakci na změny struktur sestavování výsledků vyhledávání vyhledávači. Díky této práci jsem získal mnoho prohlubujících znalostí, hlavně jazyka C#, .NET Frameworku a optimalizace webových stránek pro vyhledávače. Vytěžené znalosti a zkušenosti zajisté využiji v rozšiřování a vylepšování této aplikace či jiných projektů.

## 8 Reference

- [1] KUBÍČEK, Michal. *Velký průvodce SEO: jak dosáhnout nejlepších pozic ve vyhledávačích*. Vyd. 1. Brno: Computer Press, 2008, 318 s. ISBN 978-80-251-2195-5.
- [2] KUBÍČEK, Michal a Jan LINHART. *333 tipů a triků pro SEO: [sbírka nejlepších technik optimalizace webů pro vyhledávače]*. Vyd. 1. Brno: Computer Press, 2010, 262 s. ISBN 978-80-251-2468-0.
- [3] STATCOUNTER. *GlobalStats* [online]. 2013 [cit. 2013-03-26]. Dostupné z: <http://gs.statcounter.com/>
- [4] LINK-ASSISTANT.COM. *SEO PowerSuite* [online]. 2013 [cit. 2013-03-26]. Dostupné z: <http://www.link-assistant.com/>
- [5] WEB CEO LTD. *Web CEO* [online]. 2013 [cit. 2013-03-26]. Dostupné z: <http://www.webceo.com/>
- [6] WEBOVÝ SERVIS COMPANY S. R. O. *Seo Servis* [online]. 2008 [cit. 2013-03-26]. Dostupné z: <http://seo-servis.cz/>
- [7] SEO LINHART S.R.O. *SEO Nástroje* [online]. 2010 [cit. 2013-03-26]. Dostupné z: <http://www.seonastroje.cz/>
- [8] TECHTERMS.COM. *TechTerms.com* [online]. 2013 [cit. 2013-04-09]. Dostupné z: <http://www.techterms.com/>
- [9] MICROSOFT. *MSDN Library* [online]. 2013 [cit. 2013-04-30]. Dostupné z: <http://msdn.microsoft.com/en-us/library>
- [10] JAMES NEWTON-KING. *Json.NET* [online]. 2013 [cit. 2013-04-26]. Dostupné z: <http://james.newtonking.com/projects/json-net.aspx>
- [11] MICROSOFT. *Html Agility Pack* [online]. 2013 [cit. 2013-04-26]. Dostupné z: <http://htmlagilitypack.codeplex.com/>

## A Obsah DVD

Adresář	Popis
/VS/Download	Visual Studio 2012 Project
/App/Download	Spustitelná aplikace
/Doc	Text bakalářské práce

Tabulka 5: Obsah DVD

## B Datový Slovník

Stazeno				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Keyword	Text	NE	NE	NE
Datum	Datum a čas	NE	NE	NE
Searcher	Text	NE	NE	NE
Position	Číslo	NE	NE	NE
URL	Text	NE	NE	NE
Description	Memo	NE	NE	ANO

StazenoPaid				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Keyword	Text	NE	NE	NE
Datum	Datum a čas	NE	NE	NE
URL	Text	NE	NE	NE
Searcher	Text	NE	NE	NE

keywords				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Id.Keyword keyword vaha	Automatické číslo	ANO	NE	NE
	Text	NE	NE	NE
	Číslo	NE	NE	ANO

vlastniWeby				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Id Domena	Číslo	ANO	NE	NE
	Text	NE	NE	NE

vlastniWebyUrls				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
ID_domeny	Číslo	NE	ANO	NE
URL	Text	ANO	NE	NE

<b>konkurencniWeby</b>				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Id	Číslo	ANO	NE	NE
Domena	Text	NE	NE	NE

<b>konkurencniWebyUrls</b>				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
ID_domeny	Číslo	NE	ANO	NE
URL	Text	ANO	NE	NE

<b>storedResults</b>				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Keyword	Text	NE	NE	NE
Datum	Datum a čas	NE	NE	NE
Searcher	Text	NE	NE	NE
Position	Číslo	NE	NE	NE
URL	Text	NE	NE	NE
Description	Memo	NE	NE	ANO

<b>storedResultsPayed</b>				
Atribut	Datový typ	Primární Klíč	Cizí klíč	Null
Keyword	Text	NE	NE	NE
Datum	Datum a čas	NE	NE	NE
URL	Text	NE	NE	NE
Searcher	Text	NE	NE	NE